ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Working Memory Predicts New Word Learning Over and Above Existing Vocabulary and Nonverbal IQ

Shelley I. Gray,[a] Roy Levy,[a] Mary Alt,[b] Tiffany P. Hogan,[c] and Nelson Cowan[d]

[a] Arizona State University, Tempe [b] The University of Arizona, Tucson [c] MGH Institute of Health Professions, Boston, MA [d] University of Missouri, Columbia

ARTICLE INFO

Article History:
Received July 21, 2021
Revision received October 13, 2021
Accepted November 4, 2021

Editor-in-Chief: Stephen M. Camarata
Editor: Megan York Roberts

https://doi.org/10.1044/2021_JSLHR-21-00397

ABSTRACT

Purpose: The purpose of this study was to use an established model of working memory in children to predict an established model of word learning to determine whether working memory explained word learning variance over and above the contributions of expressive vocabulary and nonverbal IQ.
Method: One hundred sixty-seven English-speaking second graders (7- to 8-year-olds) with typical development from two states participated. They completed a comprehensive battery of working memory assessments and six word learning tasks that assessed the creation, storage, retrieval, and production of phonological and semantic representations of novel nouns and verbs and the ability to link those representations.
Results: A structural equation model with expressive vocabulary, nonverbal IQ, and three working memory factors predicting two word learning factors fit the data well. When working memory factors were entered as predictors after expressive vocabulary and nonverbal IQ, they explained 45% of the variance in the phonological word learning factor and 17% of the variance in the semantic word learning factor. Thus, working memory explained a significant amount of word learning variance over and above expressive vocabulary and nonverbal IQ.
Conclusion: Results show that working memory is a significant predictor of dynamic word learning over and above the contributions of expressive vocabulary and nonverbal IQ, suggesting that a comprehensive working memory assessment has the potential to identify sources of word learning difficulties and to tailor word learning interventions to a child's working memory strengths and weaknesses.
Supplemental Material: https://doi.org/10.23641/asha.19125911

Working memory is the active human memory process responsible for storing and manipulating incoming information. There is compelling evidence that scores on working memory measures are related to static measures of past learning (Alloway et al., 2009; Maehler & Schuchardt, 2016) and to academic achievement (Alloway & Alloway, 2010; Alloway et al., 2009; Berninger et al., 2010; Chalmers & Freeman, 2018; Gathercole et al., 2003; H. L. Swanson & Berninger, 1996). Currently, however, we lack evidence that working memory predicts variance in *dynamic* measures of learning. If we had this evidence, it could be an important step toward determining if working memory is more than just a covariate of learning, but rather is a key part of the learning mechanism. This is important to establish because if working memory has a causal relationship with dynamic learning, it might create a path for improving learning involving interventions that support working memory (e.g., by avoiding overloading working memory at any point in the learning process; see the works of Cowan, 2014; Jaroslawska et al., 2016; Sweller, 2011). If working memory is directly involved in dynamic aspects of word learning, we might expect working memory measures to account for variance in word learning even after removing variance from standardized tests of nonverbal intelligence and static measures of vocabulary, a possibility examined here.

We present data and models to predict dynamic measures of word learning in second-grade children using an extensive working memory battery. In the following sections of the introduction, to set the stage, we review

Correspondence to Shelley I. Gray: Shelley.Gray@asu.edu. **Disclosure:** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

several topics in turn. We first review models of working memory in children, as these inform the breadth of assessment necessary to fully represent the collection of working memory processes. Next, we review research investigating the relationship between different components of working memory and important aspects of cognitive development, including oral language, reading, and other areas of academic achievement, as well as limited research investigating the relationship between components of working memory and dynamic learning. Finally, we review findings from a study investigating the structure of word learning in children to establish the breadth of assessment necessary to fully represent the early stages of word learning. These topics all contribute to the rationale for this study, which is then described.

Working Memory Models

Baddeley and colleagues developed two prominent models of working memory, each influential in the area of oral language learning. The first included three distinct working memory components: central executive, phonological loop, and visuospatial sketchpad (specified by Baddeley, 1986; initiated by Baddeley & Hitch, 1974). In 2000, Baddeley added a fourth working memory component to the earlier three-component model: the episodic buffer. In the Baddeley models, the central executive is responsible for regulating attention, linking working memory to long-term memory, and directing use of other working memory components. The phonological loop is dedicated to phonological information processing, making it essential for oral and written language learning. The phonological loop has its own temporary storage capacity, which can be refreshed by verbally rehearsing material. The visuospatial sketchpad is dedicated to processing visual and spatial information, but it may lack the capacity for information rehearsal. Finally, the episodic buffer was proposed as a temporary information store capable of binding and holding information from multiple stores in relation to one another.

Cowan (1988, 1995, 1999, 2016, 2019) proposed a third prominent model of working memory, the "embedded-processes model," which focuses on a three-level hierarchy of working memory emphasizing the role of attention. These include central executive, focus of attention, and phonological storage and rehearsal factors. According to Cowan, there is the memory system as a whole and, within it, a pool of currently activated long-term memory (which can include just-learned as well as reactivated information). Furthermore, within this pool of activated long-term memory is the portion of working memory information that is in the focus of attention, allowing further processing of up to several items at once. Thus, in Cowan's historical model of working memory, there was more reliance on the role of attention in processing active information and less on separate components of working memory, which are central to Baddeley's models. Interference between items was proposed to depend on inter-item feature similarity, and the use of new binding was said to originate in the focus of attention and to result in rapid new learning that added to activated long-term memory and could be used immediately. Baddeley and Cowan agree on many issues, including central executive processes and covert phonological rehearsal. The most primary remaining disagreement concerns whether attention is needed for maintenance of phonological information, when rehearsal is not possible, and for visual information.

Gray et al. (2017) compared the fit of Baddeley (2000), Baddeley and Hitch (1974), and Cowan (1988, 1999, 2005) working memory models in second-grade children with typical development (TD). They found that the Baddeley and Hitch (1974) three-component model and the Cowan embedded-processes model each demonstrated good fit, but the Cowan model showed better fit overall. Although the Baddeley (2000) four-component model converged, the episodic buffer factor was empirically indistinguishable from the visuospatial sketchpad factor; therefore, the Baddeley (2000) four-component model was not supported. Based on loadings from the two remaining models, Gray et al. (2017) proposed a hybrid three-component model of working memory with central executive, focus of attention/visuospatial sketchpad, and phonological storage and rehearsal/phonological loop components. This model allowed cross-loadings between the visuospatial sketchpad/focus of attention and phonological storage and rehearsal/phonological loop factors. The fit for the hybrid model and Cowan's-embedded processes model was very close; therefore, we used the hybrid model in the current investigation to represent the convergence of prominent models of working memory.

Working Memory and Oral Language

Because oral language is central to most learning, many studies have examined the relationship between scores on working memory measures and scores on oral language measures. Typically, studies assess a single aspect of working memory in relation to a single component of oral language. They show that verbal working memory measures positively correlate with vocabulary and grammar scores in a person's first language (Adams & Gathercole, 1996, 2000; Baddeley et al., 1998) and second language (French & O'Brien, 2008; Masoura & Gathercole, 2005; Verhagen & Leseman, 2016). Conversely, verbal working memory deficits negatively correlate with recognition and production of words, grammatical forms, and syntactic structures (Andrade & Baddeley, 2011; Gathercole & Baddeley, 1990; Marini et al., 2014; Montgomery et al., 2010).

In the area of language comprehension, Daneman and Merikle (1996) conducted a meta-analysis of working memory and language comprehension encompassing 77 studies. They set out to resolve the paradoxical results of studies suggesting that even though short-term memory capacity should be related to reading and listening comprehension, researchers often found no correlation between short-term memory measures such as digit span and word span and scores on language comprehension measures. They argued that a working memory task (such as their own working span measure; Daneman & Carpenter, 1980) should assess combined processing and storage rather than just capacity. Their meta-analysis showed that this type of measure was more strongly correlated with language comprehension than with single-capacity measures. However, it is important to note that their working span task is language based as it requires participants to listen to or read a series of unrelated sentences and then to recall the final word of each sentence. Thus, it is not a suitable measure of working memory in participants with known language deficits because it may be a stronger indicator of oral language than working memory. Meta-analysis results showed that their processing and storage measure of working memory consistently correlated with other language measures. There have been entirely spatial analogues to the working span measure, and it has been proposed that they involve storage that is separate from the verbal tasks but that both modalities of tasks share a common factor as well, termed *executive attention* (Kane et al., 2004). The key point is that tasks that require both storage and processing of information are less likely to be targeted measures of working memory capacity and more likely to reflect contributions of multiple abilities.

Recently, Chow et al. (2021) evaluated the contribution of verbal working memory, as assessed by the Verbal Attention and Numbers Reversed subtests of the Woodcock-Johnson IV Tests of Cognitive Abilities (Schrank et al., 2014); attention, as assessed by teacher completion of the Strengths and Weaknesses of Attention-Deficit/Hyperactivity Disorder Symptoms and Normal Behavior Scale (J. Swanson et al., 2001); and static measures of child language, as measured by subtests from the Test for Auditory Comprehension of Language–Fourth Edition (TACL-4; Carrow-Woolfolk, 2014) and the Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5; Wiig et al., 2013) screening tool. Participants included 414 first and second graders, 49 of whom were English learners and 20 who received special education. Results of structural equation modeling analyses suggested that the verbal working memory measures were stronger predictors of language than attention. Results from exploratory multiple regression analyses showed that verbal attention was the only cognitive predictor of all language indicators (CELF-5 Screener; Vocabulary, Morphology, and Syntax subtests from the TACL-4). The authors concluded that when assessing working memory, it is important to consider the verbal demands of the working memory task. They also noted that their study lacked nonverbal and visuospatial measures.

Working Memory and Reading

A recent meta-analysis investigated the relationship between working memory domains, reading decoding, and reading comprehension across multiple grade levels, with 197 studies meeting inclusionary criteria (Peng et al., 2018). Domains of working memory represented in the studies included verbal, numerical, visuospatial, and composite measures. Reading measures included phonological coding, decoding, vocabulary, and comprehension. The authors reported a moderate correlation between reading and working memory overall (r = .29) with a range of .21–.35. They reported that the domain of working memory and students' grade level influenced the working memory–reading relationship. Prior to fourth grade, findings suggested that all domains of working memory related to reading similarly; however, starting in fourth grade, verbal working memory had a stronger relationship with reading than visuospatial working memory. Results of this comprehensive meta-analysis suggest there is a small but significant and consistent relationship between working memory domains and static measures of reading.

Working Memory and Academic Achievement

Previous research suggests that working memory may be a unique predictor of academic outcomes in children, over and above contributions of IQ (Alloway & Alloway, 2010; Passolunghi et al., 2007; but see the works of Giofre et al., 2017, for an exception). Alloway et al. (2009) administered measures of working memory, verbal and performance IQ, reading and spelling ability, and math to 64 children ages 7–11 years who were receiving special education for learning difficulties. Thirty-seven children were retested 2 years later on standardized reading and math measures. The authors reported that when working memory and prior scores on reading and math were controlled statistically, IQ was not a significant predictor of static learning outcomes. Alloway and Alloway (2010) administered working memory, IQ, literacy, and numeracy assessments to children in kindergarten and again 6 years later. They found that working memory in kindergarten was a more powerful predictor of later academic success than IQ.

Working Memory and Dynamic Word Learning

With a predictive relationship between working memory and static measures of oral language, reading, and academics well established, we argue that what is now needed is an understanding of the relationship between working memory and dynamic learning. That is key if we are to determine whether supports for working memory can actually improve learning in educational contexts.

Word learning experiments provide fertile ground for examining the working memory–learning relationship because word learning itself is central to language and academic success. The advantages of dynamic word learning experiments include control for prior experience and existing knowledge; reduced bias (e.g., Camilleri & Botting, 2013); and the creation of a learning situation where the child's working memory abilities, stored information in long-term memory (e.g., phonological representations), and other cognitive abilities interact dynamically.

Existing vocabulary and nonverbal IQ have each been shown to relate to vocabulary acquisition in children. For example, Verhagen and Leseman (2016) found that in learning a second language, verbal short-term memory was related to vocabulary learning, but verbal working memory was related to grammar learning. Rowe et al. (2012) showed that vocabulary growth in young children predicted later vocabulary acquisition. Finally, Rice and Hoffman (2015) showed that nonverbal IQ was a predictor of receptive vocabulary in children and young adults with and without specific language impairment (SLI) ages 2;6 (years;months) through 21.

Although Baddeley et al. (1998) summarized a great deal of work linking phonological working memory to vocabulary acquisition, few experiments have investigated relations between multiple types of working memory and word learning. In one such experiment, Morra and Camba (2009) investigated whether three constructs, phonological sensitivity, attentional capacity, and rehearsal, plus a static measure of vocabulary, predicted new paired-word learning in 8- to 10-year-olds in Italy. They found that phonological sensitivity, vocabulary, and attentional capacity related differently to short and long words that did or did not contain native phonology. The authors concluded that vocabulary learning is a "complex multidetermined phenomenon – more complex than suggested by other current models" (p. 175), suggesting that it requires new psychological models to account for this complexity.

In another experiment on multiple forms of working memory and word learning, Archibald and Joanisse (2013) compared the performance of school-age children with SLI, working memory impairment, or both disorders to children with TD on phonological short-term memory, working memory, and paired-associates word learning tasks. To attain sufficient sample sizes, they added some children with primary language impairment plus working memory impairment to the SLI group and children with primary working memory impairment plus SLI to the working memory impairment group. Groupings were based on standardized scores from nonverbal IQ and oral language tests. The word learning task included one condition that used proper nouns (real words) as names for space aliens and a second that used nonwords for space alien names. All groups learned word pairs best in the real word condition. Phonological short-term memory was linked more strongly to nonword learning, but working memory was related to both real word and nonword learning. The primarily working memory impairment group had poor learning of both words and nonwords, whereas the primarily SLI group had the most difficulty with nonword learning. The authors concluded that working memory was related to learning across modalities, but domain-specific phonological memory was related to nonword learning. These findings point to the importance of investigating different working memory factors in relation to learning to understand the nature of their relationship.

Models of Word Learning

In response to the need for a unifying theoretical model of word learning, Gray et al. (2020) tested four latent variable models encompassing the triggering and configuration stages of word learning in second-grade children with TD. In the unidimensional model, phonological, semantic, and phonological–semantic linking indicators all loaded on a single word learning factor. In the receptive/expressive model, indicators requiring children to recognize new words loaded on the receptive factor, and indicators requiring children to produce new words loaded on the expressive factor. In the phonological/semantic model, indicators assessing children's recognition or production of the phonological aspects of new words loaded on the phonological factor, and indicators assessing children's recognition or recreation of the visual (semantic) representations of new words loaded on the semantic factor. Indicators assessing the link between phonological and semantic representations loaded on both factors. Finally, in the create/recreate/link model, indicators assessing the creation and storage of new phonological and semantic representations loaded on the create factor, indicators assessing the ability to produce or recreate semantic features of referents loaded on the recreate factor, and indicators linking phonological and semantic representations loaded on the link factor. The phonological/semantic model best fit the data and, therefore, was used as the measurement model for word learning in this study.

This Study

The purpose of this study was to use an established model of working memory components in children, which included factors representing all working memory processes (Gray et al., 2017), to predict an established model of word learning (Gray et al., 2020). The working memory and word learning models were derived from analyses with the same group of second-grade children with TD who participated in this study.

The word learning experiment assessed children's ability to learn novel words during the triggering (Hoover et al., 2010) and configuration (Leach & Samuel, 2007) stages of word learning. Each of these stages illustrates how language knowledge stored in long-term memory can interact with new information being processed by working memory. Triggering occurs when a learner hears a new word form, compares it to stored word forms in long-term memory, and recognizes that the word does not match a stored form. Children who have already stored the speech sounds of their language need to recognize and remember the order of sounds in a new word; thus, their stored representations of sounds aid the creation of new phonological word forms (Norris et al., 2018).

After triggering, the phonological (individual sounds) and lexical (whole word) forms of a new word are stored in long-term memory, followed by lexical configuration that occurs with repeated exposures to the word in varying contexts. Lexical configuration comprises the incremental storage of factual information about the word in long-term memory including its meaning(s) and syntactic roles. Leach and Samuel (2007) also described the engagement stage of word learning (not included in this study) as the dynamic behavior of the new word's lexical representation in relation to other words stored in the lexicon. For example, words can activate each other when they are related phonologically, semantically, or orthographically.

This study was designed to ascertain the variance that working memory could account for in word learning over and above the contributions of existing expressive vocabulary and nonverbal IQ, as each of these has been shown to be related to vocabulary acquisition in children. We expected that nonverbal IQ would be most strongly related to the semantic aspects of word learning (represented visually in this study) because in our previous investigation of the relationship between working memory and nonverbal IQ (Gray et al., 2017), we found that the focus of attention/visuospatial sketchpad factor was strongly related to the Gv (general visual) factor of the Kaufman Assessment Battery for Children, Second Edition (KABC-II; Kaufman & Kaufman, 2004). Based on literature showing predictive relations between working memory and a variety of academic skills, our hypothesis was that working memory processes as a whole would account for variance in the phonological word learning factor (PHON) and semantic word learning factor (SEM) over and above that accounted for by extant vocabulary and nonverbal IQ.

Method

We enrolled 167 second-grade children with TD who came from the Phoenix and Tucson metropolitan areas of Arizona and the Boston metropolitan area of Massachusetts (72 girls, 95 boys). Students were recruited from multiple classrooms within multiple schools in more than 25 different districts and charter school organizations. They were part of a larger study on word learning and working memory.[1] Of the total participants consented for the larger study, 43% were from the Tucson metro area, 42% from the Phoenix metro area, and 15% from the Boston metro area.

Following institutional review board approval from our universities and local school board or organizational administrator approval, teachers, reading specialists, and speech-language pathologists serving second graders were asked to send home information packets to families. Each packet contained a cover letter explaining the study in brief and a parent consent form. Parents who wished to consent to their child's participation could return the form to their child's teacher or mail it back to us in a postage-paid envelope. We also recruited more generally in the community (e.g., through libraries, at community events like Trunk-or-Treat, and via Facebook ads). Eighty-seven percent reported their child's ethnicity as non-Hispanic, 12% reported as Hispanic, and 1% did not report ethnicity. Two percent reported American Indian or Alaska Native as their race, 2% Asian, 2% Black, 81% White, 12% more than one race, and 1% did not report. Additional information about participants is in Table 1.

To be included in this as well as in the larger study, children were required to (a) pass a bilateral hearing screening, (b) pass a color vision screening, (c) pass a near-vision acuity screening, (d) be enrolled in or have just completed second grade, (e) have no history of neuropsychiatric disorders (e.g., attention-deficit/hyperactivity disorder, autism spectrum disorder) by parent report, (f) speak monolingual English by parent report, (g) achieve a standard score $\geq$ 75 on the Nonverbal Index of the KABC-II (Kaufman & Kaufman, 2004), (h) have no history of special education services or grade repetition, (i) achieve a standard score > 30th percentile on the

[1]Participants in this study represent a portion of the participants in a larger sample from the Profiles of Working Memory and Word Learning (POWWER) project funded by National Institute on Deafness and Other Communication Disorders Grant R01 DC010784. The POWWER project includes the group reported, as well as children with SLI (now referred to as developmental language disorder), children with dyslexia, children with comorbid dyslexia and SLI, and Spanish–English bilingual children with TD. All POWWER participants completed a total of six word learning games and a comprehensive battery of working memory tasks (see the work of Cabbage et al., 2017) over the course of at least 6 days.

Table 1. Participant characteristics and test scores.

Measure	*M*	*SD*
Age in months	92.82	4.98
Mother's education in years	15.39	1.66
GFTA-2 articulation accuracy percentile	50.89	8.54
KABC II Nonverbal Index standard score	117.60	15.53
TOWRE-2 Word/Nonword standard score	109.45	8.40
CELF-4 Core Language standard score	108.75	9.58
EVT-2 standard score	112.39	10.95
WRMT-III PC standard score	108.23	9.85

Note. GFTA-2 = Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000); KABC II = Kaufman Assessment Battery for Children, Second Edition (Kaufman & Kaufman, 2004); TOWRE-2 = Test of Word Reading Efficiency–Second Edition (Torgesen et al., 2012); CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel et al., 2003); EVT-2 = Expressive Vocabulary Test–Second Edition (Williams, 2007); WRMT-III PC = Woodcock Reading Mastery Tests–Third Edition, Passage Comprehension subtest (Woodcock, 2011).

Goldman-Fristoe Test of Articulation–Second Edition (Goldman & Fristoe, 2000; unless scores below that percentile were due to consonant errors on a single sound), (j) achieve a standard score > 87 on the Core Language composite of the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel et al., 2003), and (k) achieve a second-grade composite standard score > 95 on the Test of Word Reading Efficiency–Second Edition (Torgesen et al., 2012). We also administered the Woodcock Reading Mastery Tests–Third Edition Passage Comprehension subtest (Woodcock, 2011) and the Expressive Vocabulary Test–Second Edition (EVT-2; Williams, 2007) for descriptive purposes.

General Procedure

After securing parental consent and child assent, we administered all assessments and experimental measures individually at a quiet location convenient for the family. All experimental tasks were from the Comprehensive Assessment Battery for Children–Word Learning and the Comprehensive Assessment Battery for Children–Working Memory (Gray et al., n.d.; summarized by Cabbage et al., 2017). Descriptives for TD group performance on the individual word learning measures (Gray et al., 2020) and individual working memory measures (Gray et al., 2017) may be found in these referenced publications. Internal consistency reliability estimates for the individual working memory tasks are reported in the study of Green et al. (2016).

Each task was presented as part of a computer-based, pirate-themed game that took approximately six 2-hr sessions to complete over about 2 weeks per child. One set of word learning tasks and working memory tasks were completed during each session in random order as determined by the computer. To encourage attention and provide a fun learning environment, children earned virtual coins at the end of each game that they could spend on their virtual pirate at the virtual pirate store. During the experiment, children sat in front of a touchscreen computer monitor next to a trained research assistant (RA). Both wore headsets with integrated microphones so that audio-recorded child responses could be transcribed later in the lab. Prior to working in the field, RAs were required to pass a quiz and two fidelity checks showing that they could administer and score each assessment and task correctly.

Word Learning

Nonwords

A pool of low-phonotactic-probability two- and four-syllable consonant-vowel-consonant (CVC) syllable structure nonwords were used. Their duration in milliseconds, biphone frequency, and summed biphone probability were very similar. Four 2-syllable nonwords from the pool were randomly assigned to each game (except that when word length was manipulated, two 2-syllable and two 4-syllable words were randomly assigned). No nonword had a phonological neighbor. Nonwords used as verbs were intransitive. Alt et al. (2017) provide a more detailed description of word characteristics.

Referents

A different set of colored sea monster drawings was used in each game. The monsters differed in shape, color, arm style, eye shape, and type of head covering but were similar in size.

Experimental Procedures

Five different word learning games each taught four different nonwords and took about 30 min to complete. Four games used nouns: one manipulated word length, one phonological similarity, one location of the referent (stationary vs. changing position), and one visual similarity of the referents. One game used verbs and manipulated actions. A more detailed description of each word learning game is in Table 2 and may also be found in the work of Gray et al. (2017).

The experimental procedures were the same for each of the five word learning games (see Table 2). Each game proceeded across four blocks. The first block provided two exposures (opportunities to hear the name and see the referent) to each of the four nonwords during the phonological–visual linking task. The second, third, and fourth blocks each presented 15 more exposures to the same four words. The computer randomized the order of task presentation (mispronunciation detection, naming, visual difference decision, visual feature recall) for Blocks 2, 3, and 4. Thus, by the end of the game, children had

Table 2. Description of word learning stimuli, tasks, and manipulations.

Stimuli	Process assessed	Experimental manipulation	Type of working memory assessed	Assessment task
Noun nonwords CVC-CVC two-syllable structure; no phonological neighbors (low neighborhood density); low biphone phonotactic probability (1.0039–1.009)	Create and store phonological form	2- vs. 4-syllable	Phonological loop capacity (length)	**Mispronunciation detection** (see monster, hear correct name or foil, press "yes" if correct name or "no" if incorrect name, receive immediate feedback)
		Phonologically similar vs. phonologically dissimilar words	Specificity of stored phonological representation	
	Retrieve and produce phonological form	2- vs. 4-syllable	Phonological loop capacity (length)	**Naming** (see monster, name it, positive feedback for trying)
		Phonologically similar vs. phonologically dissimilar	Specificity of stored phonological representation	
Noun referents Virtual sea monsters all the same size, but varied body shapes, colors, limb shapes, head coverings, and facial features	Create and store semantic representation	Stationary referent vs. referent changes location	Spatial memory	**Visual difference decision** (see monster, press "yes" if monster shown is an accurate depiction of one of the learned monsters or "no" if incorrect, receive immediate feedback)
		Visually similar referent vs. visually dissimilar referent	Specificity of stored semantic (visual) representation	
	Retrieve and recreate semantic representation	Stationary referent vs. referent changes location	Spatial memory	**Visual feature recall** (move correct color, eyes, arms, and head covering onto outline of monster, feedback based on number of correct selections)
		Visually similar referent vs. visually dissimilar referent	Specificity of stored semantic representation	
	Link phonological form and semantic representation	2- vs. 4-syllable	Phonological loop capacity (length)	**Phonological–visual linking** (hear monster name, choose 1 of 4 monsters, receive immediate feedback)
		Phonologically similar vs. phonologically dissimilar words	Specificity of stored phonological representation	
		Stationary referent vs. referent changes location	Spatial memory	
		Visually similar referent vs. visually dissimilar referent	Specificity of stored semantic representation	

(*table continues*)

Table 2. *(Continued).*

Stimuli	Process assessed	Experimental manipulation	Type of working memory assessed	Assessment task
Verb nonwords CVC-CVC two-syllable structure; no phonological neighbors (low neighborhood density); low biphone phonotactic probability (1.0039–1.009)	Create and store phonological form	None	Specificity of stored phonological representation	**Mispronunciation detection** (see monster perform action, hear correct command or foil, press “yes” if correct command for that action or “no” if incorrect command, receive immediate feedback)
	Retrieve and produce phonological form	None	Specificity of stored phonological representation	**Naming** (see monster perform an action, name it, positive feedback for trying)
Verb referents Single virtual sea monster with movement varied by speed, direction, nature of movement, and special effects such as glowing or pulsating	Create and store semantic representation	Four different referent actions	Spatial memory Specificity of stored semantic representation	**Visual difference decision** (see monster perform an action, press “yes” if action shown is an accurate depiction of one of the learned actions or “no” if not, receive immediate feedback)
	Retrieve and recreate semantic representation	Four different referent actions	Spatial memory Specificity of stored semantic representation	**Visual feature recall** (see an action and judge whether it is the correct speed, direction, nature of movement, and special effect, feedback based on number of correct choices)
	Link phonological form and semantic representation	All of the above	All of the above	**Phonological–visual linking** (pirate gives a command, choose 1 of 4 actions, receive immediate feedback)

Note. Reprinted from Gray et al., 2020. CVC = consonant–vowel–consonant.

heard each name spoken 47 times across the four blocks (2 + 15 + 15 + 15), they had completed five tasks per block for a total of 20 tasks (4 blocks × 5 tasks) completed by the end of each game, and they had responded to four prompts per task for a total of four possible correct responses per task.

To begin each block, pictures of four sea monsters appeared on the screen for the phonological–visual linking task. The child heard the sea monster's name or a name for the action in the case of verbs. The task was to touch the monster that corresponded to the name (or action). This task assessed their ability to link the label and referent for each of four monsters. Children received a gold coin for each correct answer and a rock for each incorrect answer. The dependent variable was the total number of correct responses across blocks.

For the mispronunciation detection task, four sea monsters appeared on the screen, one at a time. The child heard either the correct name (or action) for the monster or a foil with a different final consonant sound than the correct name. Children indicated whether the name was correct by pressing a key. They completed this task 1 time per block for a total of four decisions per monster. Their choice received immediate feedback in the form of a coin or a rock. The dependent variable was the correct number of responses adjusted for the 50% chance of guessing the correct yes/no answer. To determine the score, we calculated the proportion of correct recognition responses on trials with the correct target minus the proportion of incorrect yes responses on trials containing foils.

For the naming task, four different monsters appeared on the screen, one at a time, and the child's task was to name the monster (or action). They completed this task 1 time per block for a total of four productions for each monster. Children's responses were recorded by the computer for later transcription in the lab. Twenty percent of the responses were double-scored, with a point-to-point agreement between .90 and .93 across the different games. In this game, feedback was not tied to correct or incorrect responses; rather, children received gold coins for attempting to name the monster (action). The dependent variable was the total number of consonants produced correctly in each word per condition.

For the visual difference detection task, a familiar monster from the current game or a visually similar foil monster (e.g., color, type of head covering, or shape or multiple features varied) appeared on the screen. Children pressed a "yes" or "no" key to indicate whether the monster was the monster they had seen before, followed by feedback with a gold coin or a rock. They competed this task once for each monster for each block for a total of four decisions per monster. The dependent variable was calculated the same way as for the mispronunciation detection task described above.

For the visual feature recall task, a line drawing outline of a monster appeared on the screen next to a visual menu of semantic features that included four choices each of color, eye shapes, arms, and types of head covering. Children selected one of each feature to put on their monster, followed by a gold coin or a rock for each selected semantic feature. They completed this task 1 time per block for each of the four monsters, totaling 16 decisions for each of four monsters. The dependent variable was the percentage of features identified correctly across blocks.

Working Memory

We administered 13 different experimental working memory tasks on a touchscreen computer. The order of administration was randomized across and within research sessions. A general description of each task is included below with a more detailed description in Table 3.

Experimental Procedures

Children began the series of 13 working memory games by selecting a pirate avatar. When they completed each game, they received gold coins that they could spend on their pirate avatar at a virtual store.

A (different) pirate guide delivered instructions at the beginning of each game and showed the child how to play. After the demonstration, the child was required to pass training trials specific to each task (see Table 3). Training trials were similar to real trials but with more explanation. Children saw the pirate play the game first, often with animations to illustrate, to limit the language comprehension load. Then, children were asked to attempt the game and, for the first training trials, were given explicit feedback from the pirate about whether they were correct or incorrect. Then, they simply attempted the game. Incorrect attempts would be met with additional feedback from the pirate about why the response was incorrect. If they did not pass training, the game stopped, and they moved on to the next game. When they did play, children received no feedback on the accuracy of their responses; however, at the end of the game, they received a virtual pile of rocks and gold coins that reflected their overall performance.

Central Executive Tasks

These tasks assessed working memory using visual and auditory updating tasks that required storage and manipulation of information. To complete the tasks, children had to maintain activated memory representations while processing incoming information.

N-back auditory. This working memory updating task presented children with a sequence of stimuli after which they were asked to judge whether a stimulus was the same as or different from the preceding stimulus. Children saw a

Table 3. Description of working memory experimental tasks included in the Comprehensive Assessment Battery for Children–Working Memory (Gray et al., n.d.).

Task	Stimuli	Trial types	Number of training blocks and trials (in parentheses)	Number of trials correct to pass training	Number of trials and stimuli	Task length (min)	Dependent variable(s)
Central executive tasks							
N-back auditory	• Image of robot band • Tones	• Same • Different	1 training block: • Same (3) • Different (3)	4/6	54 (3 blocks each with 9 Same, 9 Different)	6.50	Mean accuracy for same and different trials combined
N-back visual	• Images of black squares with white dots	• Same • Different	1 training block: • Same (3) • Different (3)	4/6	54 (3 blocks each with 9 Same, 9 Different)	7.50	Mean accuracy for same and different trials combined
Number updating	• Visual presentation of numbers and operations	Not applicable	2 training blocks: • Each block (5)	5/5 each block	15 (3 blocks each with 5 trials)	7.20	Mean accuracy for all trials
Short-term phonological memory tasks							
Digit span	• Auditory recordings of digits 1–9 (except 7 because it is 2 syllables)	Span length (2–8 digits)	1 training block: • (2)	2/2	14 (2 trials at each span length of 2–8 digits)	4.50	Number of trials correct at each span length × span length then sum products
Digit span – running	• Auditory recordings of digits 1–9 (except 7 because it is 2 syllables)	Span length (7–10 digits)	3 training blocks: • Each block (3)	At least 1 correct for each of 3 blocks	12 (3 trials at each span length of 7–10 digits)	6.00	Average number of digits recalled in the correct order
Nonword repetition	• Auditory recordings of nonwords	Word length (2- to 5-syllable nonwords)	1 training block: • (3 two-syllable trials)	3 attempted	16 nonwords (4 each at 2-, 3-, 4-, and 5-syllable lengths)	3.00	Number of words repeated with correct consonants at each syllable length × syllable length then sum products
Short-term visuospatial memory tasks							
Location span	• An arrow pointing toward a location arranged in a circular pattern	Span length (2–6 locations)	3 training blocks: • 1 location (1) • 2 locations (2)	At least 1 at 1 location and 1 at 2 locations	12 (2 trials at each span length of 2–6 locations)	4.50	Correct number of trials at each span length × span length then sum products[a]
Location span – running	• An arrow pointing toward a location arranged in a circular pattern	Span length (5–8 locations)	3 training blocks: • 6 locations (1) • 7 locations (1) • 8 locations (1)	1/1 correct at each length	12 (3 trials at each span length of 5–8 locations)	7.50	Average number of locations correctly identified across all trials
Visual span	• Black polygons	Span length (1–6 polygons)	1 training block: • 1 polygon (1) • 2 polygons (2)	3/3	12 (2 trials at each span length of 1–6)	6.50	Correct number of trials at each span length × span length then sum products[a]
Visual span – running	• Black polygons	Span length (3–6 polygons)	1 training block: • 3 polygons (1) • 4 polygons (1)	1 correct at each length	12 (3 trials at each span length of 3–6 polygons)	7.00	Average number of polygons correctly identified in order across all trials

(*table continues*)

Table 3. *(Continued).*

Task	Stimuli	Trial types	Number of training blocks and trials (in parentheses)	Number of trials correct to pass training	Number of trials and stimuli	Task length (min)	Dependent variable(s)
Binding tasks							
Phonological binding span	• Auditory nonspeech sounds (e.g., mechanical noise) • Auditory recordings of nonwords	Span length (1–4)	1 training block: • 1 sound–nonword pair (1) • 2 sound–nonword pairs (1)	Attempt 2/2	20 sound–nonword pairs (2 trials each of 1–4 pairs per trial)	5.20	Correct number of trials at each span length × span length then sum products[a]
Visual–spatial binding span	• Image of a grid • Black polygons	Span length (1–6 polygons)	1 training block: • 1 polygon (1) • 2 polygons (1)	2/2	12 (2 trials at each span length of 1–6 polygons)	5.20	Correct number of trials at each span length × span length then sum products[a]
Cross-modal binding	• Black polygons • Auditory recordings of nonwords	Span length (1–6 polygons)	1 training block: • 1 nonword–polygon pair (1) • 2 nonword–polygon pairs (1)	2/2	12 (2 trials at each span length of 1–6 polygons)	6.50	Correct number of trials at each span length × span length then sum products[a]

Note. Reprinted from Gray et al., 2017, with permission. Copyright © 2017 Elsevier.

[a]This scoring method yields the total number of items within correctly recalled lists.

still picture of a robot band playing different instruments on the screen for 1,000 ms. They heard a series of tones that varied in frequency (1000, 1250, 1500, 1750, and 2000 Hz) for 1,000 ms per tone. After each tone was heard, children were asked to decide whether the new tone was the same as or different from the previous tone in the sequence by pushing keys labeled "same" and "different" on the keyboard. If they did not respond by the end of 3 s, the next stimulus was presented. Response accuracy was recorded by the computer. Our pilot work showed that children of the same age as those in this study were only able to complete 1-back tasks such as this.

N-back visual. This task proceeded like the auditory *N*-back task, except that children saw robots playing a game with patterned game pieces. Each game piece was a black square with different patterns of white dots that stayed on the screen for 1,000 ms. After each piece in a series was shown, the image left the screen and then the child was asked to judge whether the most recent patterned game piece was the same as or different from the preceding patterned game piece. If they did not respond by the end of 3 s, the next stimulus was presented. Response accuracy was recorded by the computer.

Number updating. This task assessed children's ability to maintain information in working memory and to update it when additional information was provided. This task was presented in the context of a toy factory. The child's task was to keep track of the running total of yoyos and teddy bears to manufacture. Initially, children were shown two digits to remember outlined by two black squares, one for the number of yoyos and the other for the number of teddy bears. These remained on the screen for 2,000 ms. Next, the children were shown an addition operation (e.g., +1) outlined in a red square for either the yoyos or the teddy bears, which they use to update the digit totals. The operation squares stayed on the screen for 500 ms. Finally, squares outlined in green (with yoyos and teddy bears in the background) appeared on the screen, which cued the child to report the updated running total for each type of toy. A correct response required the child to report correct running totals for both toys. If the child responded with an incorrect number but used that number from that trial onward to correctly report the running total, they scored a 0 for the initial incorrect trial response but received credit for subsequent responses in that trial that were correct relative to the single mistaken response. The next stimulus appeared on the screen 50 ms after the RA entered the child's response.

Short-Term Phonological Memory Tasks

These tasks assessed phonological short-term storage capacity using stimuli that minimized reliance on lexical or semantic knowledge. In running tasks, to reduce the ability to group items or to verbally rehearse them, the child did not know how many items they would see before being asked to respond.

Digit span. The digit span task required children to repeat strings of digits that varied in length from two to eight digits. Children were not explicitly told how many digits would be presented in the digit span task, but the number of digits presented steadily increased (e.g., two presentations of two digits, two presentations of three digits), making it predictable. Digits were presented 1 s apart. This task was presented to children in the context of playing a copycat game with a robot where the child repeated what the robot said. The robot read lists of numbers aloud in an adult male voice (digits 1–9 excluding 7 because it has two syllables) in random order. After a series of numbers was presented, the child saw a green rectangle on the screen to prompt them to verbally recall as many numbers as possible in sequence. The computer audio-recorded the child's responses, and the RA also entered them into the computer using a separate keyboard. The next span was presented after the RA keyed in the child's responses.

Digit span–running. The running task was similar to the digit span task, except the child saw sea monsters that spoke lists of numbers with spans from seven to 10 digits in length. Digits were presented 1 s apart. The procedures were the same as digit span, but the child did not know how many digits would be presented. The computer randomly presented three trials each at span of seven to 10 digits for 12 trials total. Children were asked to recall as many numbers as they could from the end of the list in forward order. For example, if the list presented was "5, 2, 8, 9, 4" and the children remembered three items, a correct response would be "8, 9, 4". The next span was presented after the RA keyed in the child's responses.

Nonword repetition. In this task, children helped the pirate build a candy bridge over a river by listening to a nonword and then repeating it. Each repeated nonword (correct or not) added one candy piece to the bridge. The 16 nonwords (four each at two-, three-, four-, and five-syllable lengths) each contained low-frequency biphones with no phonological neighbors. Words were presented in order from those with the fewest (two syllables) to those with the most (five syllables) syllables. After the child repeated each nonword, the RA pressed the advance key. The computer audio-recorded responses for later scoring in the lab. Children scored 1 point for each nonword with all consonants repeated correctly. We elected to score only consonants to achieve higher interrater reliability than would be possible if vowels were also scored (e.g., van Haaften et al., 2019). In addition, although RAs transcribed the children's production of consonants in nonwords for scoring, we chose to score each whole nonword production as correct or incorrect because many future users of this task may not have trained assistants to transcribe productions phonetically. Twenty percent of all transcriptions were double-scored with an interrater transcription reliability for phoneme-by-phoneme consonant scoring of 87%. If there were any discrepancies, the primary coder's transcription was included in the analysis.

Short-Term Visuospatial Memory Tasks

These tasks assessed children's short-term memory for shapes and locations that could not easily be remembered using verbal labels. In the running tasks, the child did not know how many items would be presented.

Location span. This task required children to recall a series of points displayed in sequence on the computer screen in a game to help the pirate locate buried treasure. Children saw a black dot in the center of the screen, followed by a series of arrows pointing to discrete locations radiating out from the black dot. Each arrow appeared on the screen for 1,000 ms and then disappeared, followed by the next arrow 1.3 s later. After the sequence of arrows was presented, eight red dots appeared in a circular pattern around the screen to show all the possible locations where arrows could point. The locations did not correspond to those seen on a clock face. Children were asked to touch the red dots to represent the locations of the sequence of arrows they had seen. The game advanced after the expected number of locations was touched by the child.

Location span–running. As in the location span task, children saw a black dot in the center of the screen followed by arrows pointing to locations, but they did not know how many locations they would need to remember. Each arrow appeared on the screen for 1,000 ms and then disappeared, followed by the next arrow 1.3 s later. The computer randomly presented three trials each at spans of three to six locations for 12 trials total. At the end of the sequence, eight red dots appeared in a circular pattern around the screen, and children were asked to touch the red dots to represent as many locations as they could remember from the end of the list in forward order. When they were finished, they touched a "NEXT" button to indicate they were ready to begin the next trial.

Visual span. This task was similar to the location span task except children saw a black polygon shape representing a gem appearing in the center of the screen one at a time for 1,000 ms. Polygons were spaced 2 s apart. We purposefully selected polygon shapes that would be difficult to name. When all polygons in the series had been shown, a selection screen appeared with empty response boxes equivalent to the number of polygons in the sequence. From a field of six available polygons, children selected polygons they had seen in the order in which they had appeared. When they were finished, the next trial began.

Visual span–running. As in the visual span task, children saw black polygon shapes in the center of the screen one at a time for 1,000 ms spaced 2 s apart, but they did not know how many would appear in the series. The computer randomly presented three trials each at a span of three to six polygons for 12 trials total. At the end of the series, children were prompted to recall the polygons in order by choosing from the six polygons displayed on the screen. When they were finished, they touched a "NEXT" button to indicate they were ready to begin the next trial.

Binding Tasks

For binding tasks, two different types of stimuli were presented from the phonological and/or the visuospatial domains, and children were asked to bind this information within working memory to respond to the prompt correctly.

Phonological binding span. The task was to remember pairings of auditory sounds and spoken nonwords in the context of a robot speaking a robot language ordering candy at a candy store. First, children saw a robot on the screen and heard the nonspeech sound (e.g., beep, mechanical noise) emitted by the robot for 500 ms. This was followed 2,000 ms later by a speaker icon on the center of the next screen presenting the nonword naming the candy that the robot ordered using the robot sound. A green rectangle appeared on the next screen, prompting the child to say the nonword paired with the previous nonspeech sound. For example, for a span of two, children heard one sound followed by a nonword and then another sound followed by a nonword. Then, they heard each sound one at a time and were asked to repeat each nonword as they heard each sound. By asking children to repeat the correct nonword (i.e., match it to the proper nonspeech sound), we were able to assess whether they successfully bound the nonword and the nonspeech sound together. After the child said the nonword, an RA advanced the program to the next trial.

The candy names were 11 single-syllable CVC nonwords with low phonotactic probability (seven to 13 neighbors each), which were drawn randomly by the computer for each trial. No sound or nonword was repeated within a trial. Children heard from one to four pairings in a trial. The computer audio-recorded the child's responses for later scoring in the lab. A nonword was scored correct if all consonant sounds were produced correctly. Consistent articulatory substitutions were not counted as incorrect. Interrater transcription reliability was 94%.

Visual–spatial binding span. This task required children to remember one paired visual-and-spatial piece of information at a time. Children saw a 4 × 4 grid with 16 squares on the screen. One polygon at a time appeared at a location in the grid for 1,000 ms and then the screen went blank for 500 ms, followed 1,000 ms later by the next polygon. There were 12 trials (two each at span lengths of one to six polygons). To respond, children selected polygons from a field of six that appeared on the screen next to the grid. They were asked to drag the polygons to their correct location in the order they originally appeared. The game advanced after the child pushed the "NEXT" button to indicate they were finished with their selections.

Cross-modal binding. This task required children to bind auditory and visual information via the pairing of single-syllable nonwords with black polygons. Children saw a black polygon on the screen for 1,000 ms and heard a nonword name for the polygon. From one to six, polygons were presented 2,000 ms apart in a single series. At the end of the series, a field of the six polygons appeared on the screen. As children heard a nonword name spoken, they were to touch the correct polygon on the screen. After they provided a response, the game advanced. The order of presentation on the test screen differed from the original presentation order. The nonwords used to name the polygons were dissimilar from each other. They did not contain the same vowels, and each had low phonotactic probability and neighborhood density.

Analytic Approach

Building off existing factor-analytic work on the structure of working memory (Gray et al., 2017) and word learning (Gray et al., 2020) factors, we pursued the relationships among working memory and word learning constructs via structural equation modeling (Bollen, 1989). The correlations among the working memory and word learning variables are presented in Supplemental Material S1, owing to the large size of the table.

Beginning with working memory, Gray et al. (2017) found support for a three-factor model based on a combination of the three-component model from Baddeley and Hitch (1974) and Cowan's embedded process model (Cowan, 1988, 1995, 1999, 2001, 2005). The resulting model specifies three latent factors: central executive, focus of attention/visuospatial sketchpad, and phonological storage and rehearsal/phonological loop (see Figure 1a). Turning to word learning, Gray et al. (2020) found support for a model with two latent factors: phonological (see Figure 1b) and semantic (see Figure 1c).

In the present work, we specified a structural equation model in which the factor-analytic models for the working memory factors and the word learning factors served as measurement models. The predictors of the two word learning factors included the (a) three working memory factors and (b) EVT-2 (expressive vocabulary) and KABC-II (nonverbal IQ) standard scores.[2] The EVT-2 scores were modeled as a single indicator of a factor, and KABC-II nonverbal scores were modeled as a single indicator of a factor. This approach distinguishes between the observed scores and the theoretical underlying latent construct via the use of estimates of the reliability of the scores (Kline, 2015). Based on their test manuals, the estimated reliability for the EVT-2 scores is .92 (Williams, 2007), and the estimated reliability for the KABC-II nonverbal scale scores is .95 (Kaufman & Kaufman, 2004).

The word learning factors were also permitted to be correlated above and beyond that due to their dependence on these predictors, as there is no theoretical reason to hypothesize that these predictors would fully account for the association between the word learning factors. The model is depicted in Figures 1 and 2. Figures 1a–1c depict the measurement models for working memory, phonological word learning, and semantic word learning, respectively. Figure 2a depicts the structural model for the latent predictors of word learning, and Figure 2b depicts the structural model relating working memory, as well as EVT-2, and KABC-II scores to word learning.

We conducted several analyses to pursue the hypothesis that working memory would account for a significant amount of variance in phonological and semantic word learning over and above the variance accounted for by expressive vocabulary and nonverbal IQ. In addition to the just-described model, a second model was fit. The second model differed from the first by constraining the effects of the working memory factors on the word learning factors to be 0. The difference in the amount of variance in word learning factors that are explained by these models is interpreted as the variance that can be attributed to the working memory factors above the EVT-2 and KABC-II nonverbal scores. We also conducted several other analyses but defer their description until presentation of the results from the models described so far.

The models were fit using the lavaan package (Rosseel, 2012) in R. We employed full information maximum likelihood to accommodate the missing data and obtained robust standard errors and fit statistics to better accommodate departures from normality (Satorra & Bentler, 1994). As evidenced by the diagonal in Supplemental Material S1, there were only a few missing cases for most variables. These were due to technology failures and, as such, are treated as missing completely at random.

Following convention, each model was evaluated by considering the model χ^2 statistic, which tests the null hypothesis of correct model specification. As this hypothesis is known to be false a priori, the model was also evaluated using fit indices, including the comparative fit index (CFI; Bentler, 1990), the standardized root-mean-square residual (SRMR; Bentler, 2006), and the root-mean-square error of approximation (RMSEA; Steiger & Lind, 1980; see the work of Steiger, 2016), the last of which also supports a hypothesis test of close fit (RMSEA $\leq$ 0.05)

[2]The models described were initially fitted using the KABC-II scores. However, the fitted models exhibited numerical instability, possibly due to the large variance associated with the KABC-II scores. These scores were rescaled by dividing them by 10 before fitting the models; no evidence of numerical instability was found for any of the models using the rescaled KABC-II scores.

Figure 1. Path diagram with standardized estimates for the measurement portions of the structural equation model listed as Model 1 in Table 4, including the (a) measurement model for the working memory factors, (b) measurement model for the word learning phonological factor (Phonological), and (c) measurement model for the word learning semantic factor (Semantic). [ns] indicates the path was not significant at the .05 level. Variances for the errors for the observed indicators and the disturbances for the endogenous latent variables are represented with one-headed arrows without a source; the associated estimated error or disturbance variance is printed by the arrow.

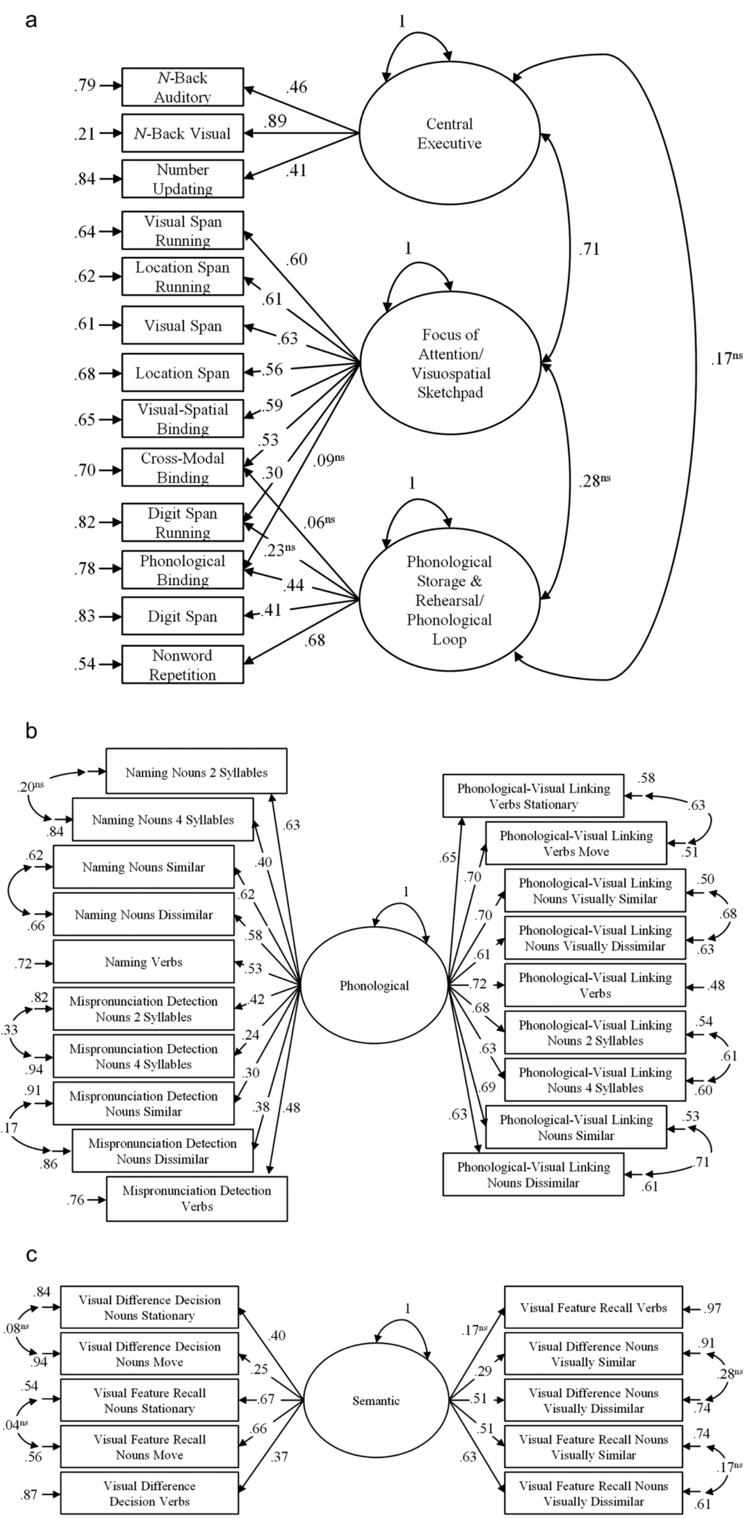
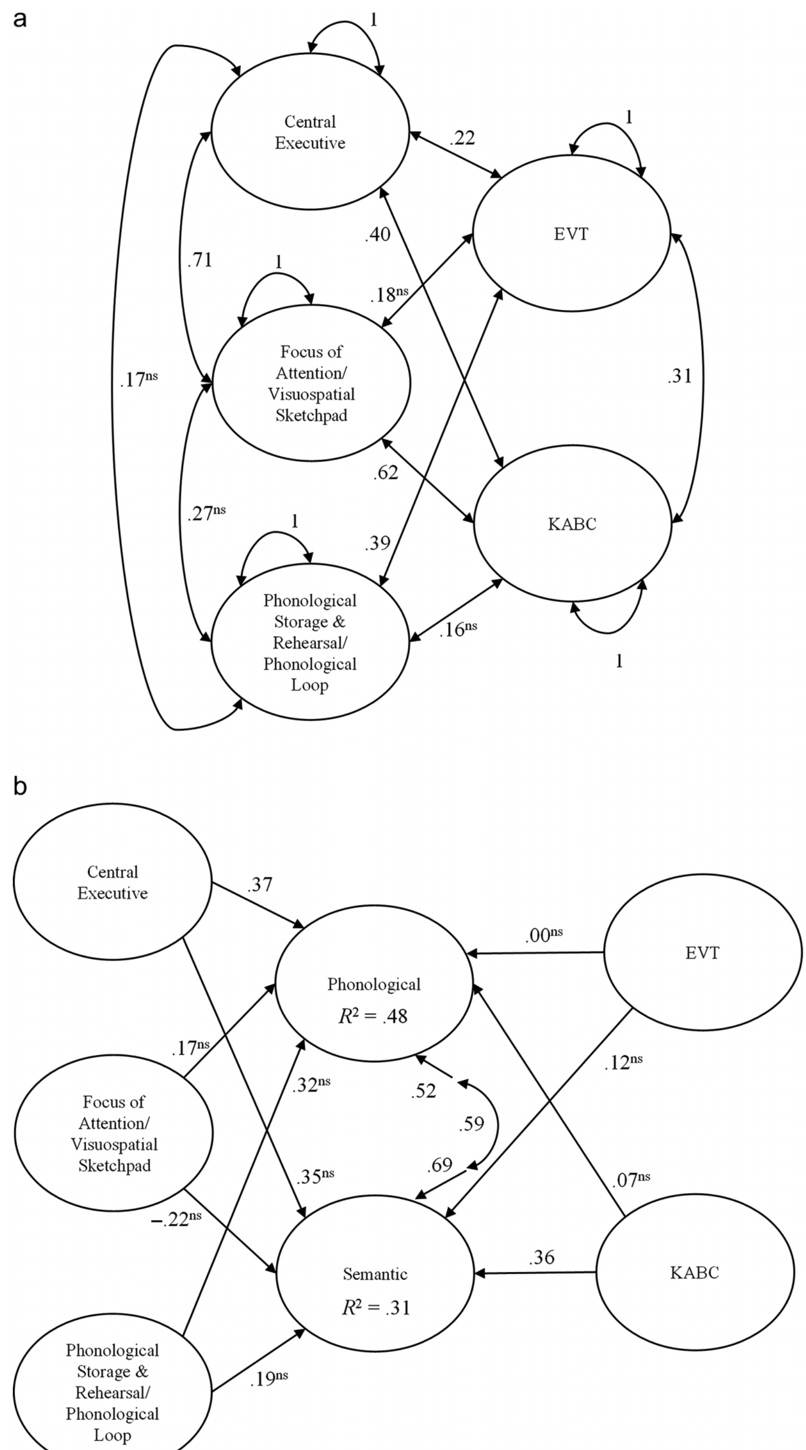


Figure 2. Path diagram with standardized estimates for the structural equation model listed as Model 1 in Table 4, including the (a) covariance structure among the latent predictors and (b) structural model for the latent predictors predicting the word learning factors (Phonological = phonological word learning factor, Semantic = semantic word learning factor). Note that the (co)variances among the latent predictors are not shown in (b); they are shown in (a). The measurement component is not shown. For the working memory factors, the structure of the measurement model was depicted in Figures 1a–c. For the Expressive Vocabulary Test (EVT) and Kaufman Assessment Battery for Children (KABC) factors, single indicators were used in the form of Expressive Vocabulary Test–Second Edition and Kaufman Assessment Battery for Children, Second Edition standard scores. [ns] indicates the path was not significant at the .05 level. Variances for the disturbances for the endogenous latent variables are represented with one-headed arrows without a source; the associated estimated error or disturbance variance is printed by the arrow.



and a confidence interval. The difference in fit between the two models was evaluated via a likelihood ratio (χ^2 difference) statistic.

In addition, we leveraged the latent variable modeling framework to characterize reliability. Evaluating the reliabilities for the observed scores from the working memory and word learning tasks separately (i.e., one task at a time) is sensible if each such task would be used in isolation. In the present work, we model the tasks as indicators of constructs represented by latent factors. We therefore seek to understand the reliability for sets of tasks. More specifically, we evaluate the maximal reliability for the set of tasks with respect to a factor (Bentler, 2007; Hancock & Mueller, 2001; Raykov, 2004) via Coefficient *H,* which is a function of the standardized loadings for the indicators of a factor (Hancock & Mueller, 2001). This conforms to the use of factors to represent constructs as opposed to, say, unit-weighted composites (McNeish, 2018). In this work, we report Coefficient *H* for each factor based on its indicators, using the Excel spreadsheet provided by McNeish (2018).

Results

A summary of the fit of Model 1 (linking working memory factors, word learning factors, and standardized tests) is given in the first row of Table 4. Though the model χ^2 statistic was statistically significant, $\chi^2(868)$ = 1,073.349, $p < .001$, the results for the CFI (.911), SRMR (.065), and RMSEA (0.037, 90% CI [0.029, 0.044]) are indicative of close model–data fit (Bentler, 1990; Browne & Cudeck, 1993; Hu & Bentler, 1999). Figure 1 depicts the *measurement* portion of the model, essentially describing the components of working memory and word learning, with standardized estimates for the model parameters. Figure 2 depicts the *structural* portion of the model, describing the relation between working memory factors, standardized test scores, and word learning factors, with standardized estimates for the model parameters. Complete unstandardized and standardized results for all parameters are given in Supplemental Material S2–S8.

Beginning with the measurement model components, the results for the measurement structure of the working memory factors mimicked those reported by Gray et al. (2017). Likewise, the results for the measurement structure of the word learning factors mimicked those reported by Gray et al. (2019).

Model 1 (see Figure 2b) explained 48% of the variance in phonological word learning and 31% of the variance in semantic word learning. A model restricting the effects of the working memory factors on word learning factors to 0 (see Table 4, Model 2) exhibited worse fit, where it can be seen that the χ^2 difference (based on six degrees of freedom) is equal to 48.878, $p < .001$. The Akaike information criterion, Bayesian information criterion (BIC), and sample-size adjusted BIC also point to worse fit of this model, as do the individual model fit statistics. In this model (not shown in a figure), because the effects of working memory factors were restricted to 0, the only explanatory

Table 4. Summary of fit of structural equation models predicting the word learning factors.

Model number and predictors	Model χ^2	χ^2 difference test[a]	CFI	SRMR	RMSEA [90% CI]	AIC	BIC	*n*-adjusted BIC
1. Vocabulary, nonverbal IQ, & working memory factors	$\chi^2(868)$ = 1,073.349, $p < .001$	—	.911	.065	0.037 [0.029, 0.044]	6501.458	7019.045	6493.464
2. Vocabulary, nonverbal IQ, & working memory, effects of working memory factors removed	$\chi^2(874)$ = 1,122.227, $p < .001$	$\chi^2(6)$ = 48.878, $p < .001$	.892	.088	0.041 [0.034, 0.048]	6537.216	7036.095	6529.511
Incremental SEM approach								
3. Block 1: vocabulary & nonverbal IQ Block 2: working memory	$\chi^2(874)$ = 1,092.064, $p < .001$	—	.905	.067	0.039 [0.031, 0.046]	6507.456	7006.335	6499.751
4. Block 1: working memory Block 2: vocabulary & nonverbal IQ	$\chi^2(875)$ = 1,096.856, $p < .001$	—	.904	.067	0.039 [0.031, 0.046]	6509.904	7005.665	6502.248

Note. CFI = comparative fit index; SRMR = standardized root-mean-square residual; RMSEA = root-mean-square error of approximation; CI = confidence interval; AIC = Akaike information criterion; BIC = Bayesian information criterion; SEM = structural equation modeling.

[a]Comparing the model in this row to the more general model listed in the top row of this section of the table; the test was conducted taking into account that the model χ^2 was based on maximum likelihood robust estimation (Satorra & Bentler, 2001).

work was done by the factors for the EVT-2 and KABC-II; these explained 17% of the variance in phonological word learning and 23% of the variance in semantic word learning.

By subtracting the variance explained in Model 2 for each word learning factor from that explained by Model 1 where working memory factors were not set to 0, we found that the working memory factors explained an additional 31% (48%−17%) of the variance in phonological word learning and an additional 8% (31%−23%) of the variance in semantic word learning. It can be difficult to evaluate whether this additional amount of variance explained should be considered to be statistically significant. The difference in fit between the models was statistically significant (see Table 4), which is suggestive that including the paths from the working memory factors is beneficial. However, such a test is about the overall fit of the model and is not targeted to the variance explained.

Another approach might examine the structural coefficients from the working memory factors to the word learning factors. Though these express the unique contributions of each working memory factor as a predictor, tests of them do not directly address the contribution of the set of predictors in terms of the variance explained. In regression, statistical tests of individual predictors can yield different conclusions than tests of the variance explained by a set of predictors. For example, it is possible for all the tests of the individual predictors to be nonsignificant, but a test of the variance explained by the set of predictors to be significant (Cohen et al., 2003, p. 90). It stands to reason that the same phenomenon could occur at the latent level. To illustrate in the present case, the analyses so far indicate that the working memory factors explain an additional 7% of the variance in semantic word learning above and beyond that explained by the EVT-2 and KABC-II nonverbal scores, but none of the coefficients for the working memory factors predicting semantic word learning were statistically significant (see Figure 2). It would be useful to have a procedure to more directly test the statistical significance of the additional variance explained by this set of latent predictors, akin to what occurs with tests of additional variance explained in hierarchical regression (see, e.g., Cohen et al., 2003, Section 5.5).

Incremental SEM Approach

Recently, Feng and Hancock (2021) developed an incremental approach that enacts such a test of the significance of additional variance accounted for within structural equation models. In their approach, analogous to what occurs in hierarchical regression, measured or latent variable predictors are grouped into blocks, such that the model yields the proportion of variance explained by each block of predictors, above and beyond all previously entered blocks. To pursue our hypothesis, we employed the approach of Feng and Hancock (2021), analyzing a model in which the factors for the EVT-2 (vocabulary) and KABC-II (nonverbal IQ) scores formed the first block and the working memory factors formed the second block. The structural portion of this model is displayed in Figure 3. The measurement models for these factors are the same as those in the previous model (single-indicator models for EVT-2 and KABC-II; see Figures 1a–1c for the remaining factors and those that are not depicted here).

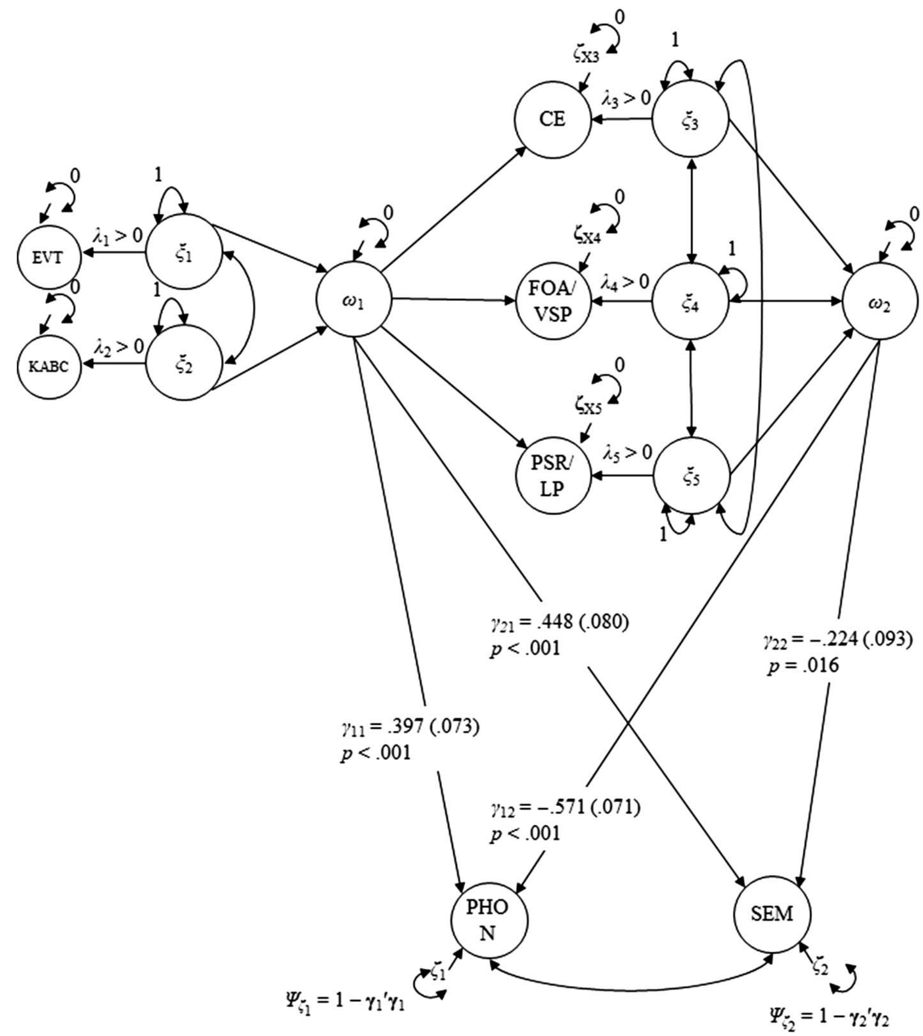
The diagram in Figure 3 portrays the structural portion of the model, along with the necessary constraints to enact a desired partitioning of variance. Toward the bottom portion of the figure, estimates of key paths (explained in more detail below) are reported. Before discussing the results, we briefly provide a conceptual presentation of the model. Technical details on the constraints and statistical theory can be found in the work of Feng and Hancock (2021).

Beginning on the left-hand side of the figure, the model essentially combines the variables in the first block (EVT and KABC) into a composite, ω_1. The use of the latent variables ξ_1 and ξ_2 and the depicted constraints on the parameters are done so that the resulting composite (ω_1) explains as much variance as possible in the outcomes (PHON and SEM).

Turning to the right-hand side of the figure, a similar structure appears for the working memory factors in the second block (central executive factor [CE], focus of attention/visuospatial sketchpad factor [FOA/VSP], and phonological storage and rehearsal/phonological loop factor [PSR/PL]). The latent variables ξ_3, ξ_4, and ξ_5 and the depicted constraints on the parameters are included to produce the resulting composite, ω_2. Importantly, note that the working memory factors in the second block (CE, FOA/VSP, and PSR/LP) are regressed on the composite from the *first* block (ω_1). In essence, this construction partitions the variance of the factors in the second block into (a) a portion that is shared with those from the first that can explain variance in the outcomes and (b) a portion that is not shared with those from the first block. The result is that the composite formed in the second block (ω_2) is that which explains as much variance as possible in the outcomes (PHON and SEM) above and beyond that which can be explained by the first block.

In summary, this modeling approach aims to partition the variance explained by the predictors into distinct portions organized incrementally by blocks (Feng & Hancock, 2021). The goal is to characterize and provide a test of the variance explained by a later block of predictors above and beyond an earlier block. The relationships specified in the reparameterization are not intended to be a theoretical model, and the resulting parameter estimates are not of inferential interest (e.g., negative γ values shown in Figures 3 and 4 are not like negative correlations), with the exception of those described next.

Figure 3. Path diagram, with notation from Feng and Hancock (2021), of the structural portion of a model for characterizing the incremental validity of the working memory factors above and beyond expressive vocabulary and nonverbal IQ. Constraints on parameters necessary to implement the model are depicted. The model is not intended as a theoretical model; it serves to partition the variance explained in the outcomes (phonological word learning factor [PHON] and semantic word learning factor [SEM]) via factors in the first block (Expressive Vocabulary Test, factor for expressive vocabulary [EVT] and Kaufman Assessment Battery for Children, factor for nonverbal IQ [KABC]) and factors in the second block (central executive factor [CE], focus of attention/visuospatial sketchpad factor [FOA/VSP], and phonological storage and rehearsal/phonological loop factor [PSR/LP]). Parameter estimates are given for the γs that are paths from composites from each block to the outcomes. Standard errors are given in parentheses, along with *p* values. The square of each of these values represents the proportion of variance explained in the outcome by the block in question, above and beyond any previous block.



The squared values of the paths from ω_1 (the composite of the first block of predictors) and ω_2 (the composite of the second block of predictors) to the outcomes give the proportion of variance that is explained by each block, in the latter case above and beyond that which is explained in the first block. That is,

- $(\gamma_{11})^2$ is the proportion of variance in the PHON that is explained by the factors for EVT-2 and KABC-II;
- $(\gamma_{21})^2$ is the proportion of variance in the SEM that is explained by the factors for EVT-2 and KABC-II;
- $(\gamma_{12})^2$ is the additional proportion of variance in the PHON that is explained by the working memory factors above and beyond that which is explained by the factors for EVT-2 and KABC-II; and
- $(\gamma_{22})^2$ is the additional proportion of variance in the SEM that is explained by the working memory factors above and beyond that which is explained by the factors for EVT-2 and KABC-II.

The model was fit in Mplus (Version 8.3; (Muthén & Muthén, 1998), adapting code provided by Feng and Hancock (2021). The fit of the model (Model 3) is summarized in Table 4. The model fit is close to that of the conventional model (Model 1 in Table 4). The estimates for the relevant parameters are given in Figure 3, in the lower portion of the figure. According to this analysis, the first block of EVT-2 and KABC-II explained about 16% of the variance in phonological word learning, as $(\gamma_{11})^2 = .397^2 = .157$, which was statistically significant, $p < .001$. The first block of EVT-2 and KABC-II explained about 20% of the variance in semantic word learning, as $(\gamma_{21})^2 = .448^2 = .200$, which was statistically significant, $p < .001$. The working memory factors explained an additional 33% of the variance in phonological word learning, as $(\gamma_{12})^2 = -.571^2 = .326$, which was statistically significant, $p < .001$. They explained an additional 5% of the variance in semantic word learning, as $(\gamma_{22})^2 = -.224^2 = .050$, which was statistically significant, $p = .016$. These values echo what was evaluated by the difference in variance explained from fitting separate models.[3]

To more fully understand the contributions of working memory as a whole as possibly distinct from expressive vocabulary and nonverbal IQ, we also examined a model (Model 4) in which the order of the blocks was reversed, with the working memory factors forming the first block and the EVT-2 and KABC-II scores forming the second block. The structural portion of this model is displayed in Figure 4, and the model fit is summarized in Table 4. The model fit is close to that of the conventional model (Model 1) including all the predictors and the previous model (Model 3) employing the incremental approach. The differences in fit arise due to the different parameters estimated under the alternate parameterization of the latent structure.

By themselves, the working memory factors were estimated as explaining about 45% of the variance in phonological word learning, as $(\gamma_{11})^2 = .673^2 = .453$, which was statistically significant, $p < .001$. The working memory factors explained about 17% of the variance in semantic word learning, as $(\gamma_{21})^2 = .408^2 = .166$, which was statistically significant, $p < .001$. Entered after these working memory factors, EVT-2 and KABC-II, as a block, explained only .1% additional variance in phonological word learning, as $(\gamma_{12})^2 = -.033^2 = .001$, which was not statistically significant, $p = .692$, and 9% additional variance in semantic word learning, as $(\gamma_{22})^2 = -.296^2 = .087$, which was statistically significant, $p = .001$.

Taking these models together, we can interpret the results and the implications for understanding the unique predictive capacity of the working memory factors as follows. Figure 5 depicts the situation via an Euler diagram in which the shapes represent the proportions of variance in the word learning factors that are explainable by the two blocks (working memory factors; vocabulary and nonverbal IQ factors). In the plots, the area of the objects corresponds to the proportion of variance, which is also labeled numerically, and the overlap between objects represents variance shared between them.

Three key aspects are readily seen in the plots. First, the predictor variables explain more variance in the PHON than in the SEM.

Second, for predicting the PHON, if the researcher only had access to vocabulary (EVT-2) and nonverbal IQ (KABC-II) scores (see Table 4, Model 3), the model would explain only 15% of the variance. Adding the working memory factors later would explain an additional 33% of the variance. However, as shown by Model 4 (see Table 4), the working memory factors can explain that same 48% of the variance by themselves, with vocabulary and nonverbal IQ explaining no additional variance. Essentially, the predictive capabilities of the working memory factors include the predictive capabilities of EVT-2 and KABC-II, and then some.

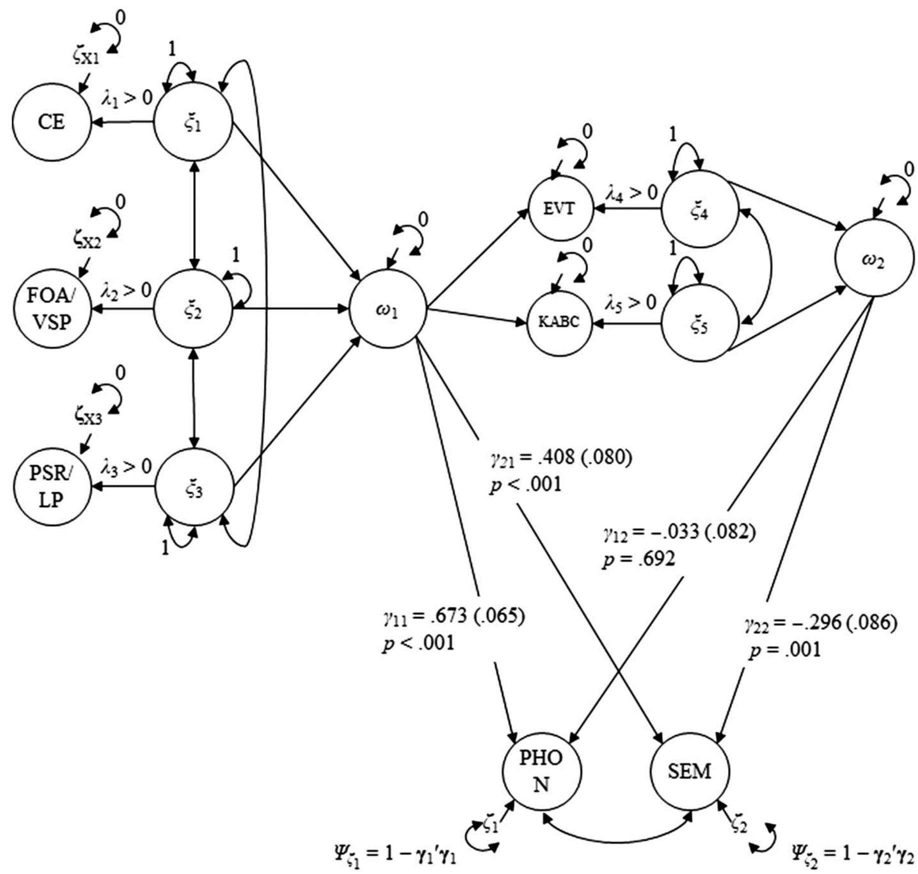
Third, the pattern for predicting the SEM is different than that for the PHON. While there is some common variance that either the working memory factors or vocabulary and nonverbal IQ scores could explain, each set of predictors can explain some variance that the other set cannot. In terms of how much variance, vocabulary and nonverbal IQ can explain about 20% of the variance by themselves. Adding working memory later would explain an additional 5% of the variance (see Table 4, Model 3). Likewise, the working memory factors can explain 16% of the variance by themselves; adding the vocabulary and nonverbal IQ factors can explain an additional 9% of the variance (see Table 4, Model 4).

Discussion

In this study, we used an established model of working memory in children to predict an established model of dynamic word learning to determine whether working memory processes as a whole explained word learning variance over and above the contributions of expressive vocabulary and nonverbal IQ. Strengths of this study include the comprehensive nature of the working memory and word learning tasks and the use of structural equation

[3]We note that the total variance explained in semantic word learning according to this analysis was about 25%, which is a bit lower than based on the conventional model, which, as noted above, was about 31%. It is not clear if this difference is due to differences in rounding or numerical imprecision under the Feng and Hancock (2021) approach, the difference in software, or some other reason. To investigate the possibility of differences due to software, all the models previously described as being fit in lavaan were also fit in Mplus, replicating the results almost exactly. As such, our conjecture is that the issue is not due to the shift in software to Mplus. Note that in the case of phonological word learning, the total explained variance was the same across modeling approaches and software (48%).

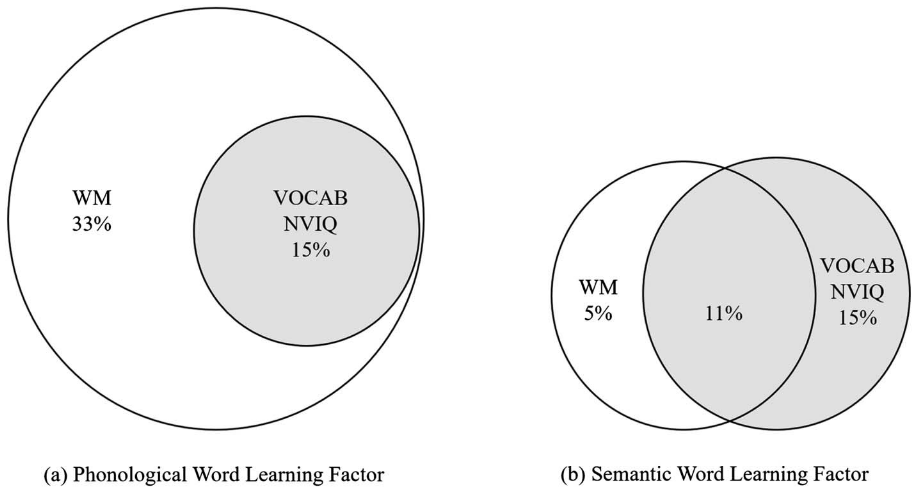
Figure 4. Path diagram, with notation from Feng and Hancock (2021), of the structural portion of a model for characterizing the incremental validity of expressive vocabulary and nonverbal IQ above and beyond the working memory factors. Constraints on parameters necessary to implement the model are depicted. The model is not intended as a theoretical model; it serves to partition the variance explained in the outcomes (phonological word learning factor [PHON] and semantic word learning factor [SEM]) via factors in the first block (Expressive Vocabulary Test, factor for expressive vocabulary [EVT] and Kaufman Assessment Battery for Children, factor for nonverbal IQ [KABC]) and factors in the second block (central executive factor [CE], focus of attention/visuospatial sketchpad factor [FOA/VSP], and phonological storage and rehearsal/phonological loop factor [PSR/LP]). Parameter estimates are given for the γs that are paths from composites from each block to the outcomes. Standard errors are given in parentheses, along with *p* values. The square of each of these values represents the proportion of variance explained in the outcome by the block in question, above and beyond any previous block.



modeling to model measurement error and assess the relations among latent working memory and word learning variables. By establishing these factor-analytic models in young children with TD, we can later test for measurement invariance of the models in other groups of children with developmental disorders who often have difficulty with word learning to determine if comparing the groups on these constructs is valid.

Results of this study support the hypothesis that working memory as a whole is a significant predictor of not only what has already been learned (academic achievement) but also what is actively being learned (dynamic learning). This is important because a logical next step would be to test the hypothesis that learning could be improved if working memory processes were optimized. This does not necessarily mean increasing working memory capacity through training, the efficacy of which has been called into question (e.g., Apter, 2012; Cunningham & Sood, 2018; Melby-Lervåg & Hulme, 2013; Randall & Tyldesley, 2016; Redick et al., 2013; Shipstead et al., 2012), although some studies show transfer effects (e.g., Holmes & Gathercole, 2014; Loosli et al., 2012; Söderqvist & Nutley, 2015). Rather, it suggests testing tailored teaching strategies to support children with particular working memory profiles (Cowan, 2014; Gray et al., 2019). This type of intervention study has yet to appear in the research literature; however, clinical focus articles such as those by Singer and Bashir (2018) provide a framework to guide development of such interventions. In addition to Singer and Bashir, there are studies showing that different manipulations of encoding practices, such as repeated and spaced retrieval (Leonard & Deevy, 2020) and effortful retrieval (Fazio & Marsh, 2019), may benefit recall and retention in children. This type of tailored teaching,

Figure 5. Euler diagrams using the working memory (WM) factors and vocabulary (VOCAB; Expressive Vocabulary Test–Second Edition scores) and nonverbal IQ (NVIQ; Kaufman Assessment Battery for Children, Second Edition standard scores) factors to explain the variance in the (a) phonological word learning factor and (b) semantic word learning factor. The areas are proportional to the percentage of the variance explained by each source, which are also labeled numerically.



combined with knowledge of a child's working memory strengths and weaknesses, has the potential to individualize teaching and intervention approaches.

An earlier study (Gray et al., 2019) found that children's working memory profiles were not synonymous with learning disability diagnoses. In that study, children with dyslexia were represented in four different working memory profiles ranging from low overall to high overall. The same was true of children with developmental language disorder, developmental language disorder and dyslexia, and TD. This suggests the potential for results from a comprehensive working memory battery to yield important information about potential ways to improve learning in addition to information already available from a typical psychoeducational evaluation that typically includes IQ and vocabulary scores. Using word learning as a dynamic learning measure, the working memory factors in this study explained a significant amount of dynamic learning variance. Quite remarkably, factors underlying the working memory battery alone accounted for nearly half of the variance in phonological word learning and nearly one fifth of the variance in semantic word learning. Adding the nonverbal IQ and expressive vocabulary measures explained no additional phonological word learning variance but an additional 9% of semantic word learning variance. Taken together, these results suggest that a comprehensive working memory assessment reflecting all components of working memory could provide important diagnostic information regarding the source of dynamic word learning difficulties.

It is important to note that structural equation modeling offers several advantages previously discussed, but such models cannot definitively pin down causation or thoroughly represent the complex working memory and word learning processes occurring in the real world. In particular, we employed models to understand the predictive utility of working memory for word learning in addition to expressive vocabulary and nonverbal IQ. A variety of alternative model structures, which impose a different directional or causal structure order on the variables, may hold. Such statistical models should be grounded in theory, and future work may serve to compare them. Ideally, this would proceed with larger samples and/or longitudinal samples. Larger samples may shed additional light on the effects, as well as the differences between effects, investigated here. Similarly, although constructs such as working memory and receptive vocabulary are represented by factors in our models, this does not mean that the skills and abilities represented by these constructs do not interact dynamically during learning.

Finally, just as structural equation models represent constructs imperfectly, tests and experimental tasks are not pure measures of a construct. For example, our measure of expressive vocabulary cannot represent a child's entire lexicon, and our nonverbal IQ measures, even though they are labeled nonverbal, do not prevent children from using oral language to help complete nonverbal tasks. Rather, they are designed to minimize reliance on oral language.

Limitations

One important limitation of this study is that we did not have the school names for each child enrolled in our study. This means that we could not determine whether our data were nested, which would permit us to test the statistical assumption of whether cases were truly independent. Because we recruited 167 participants from more than 25 different school districts and educational organizations in three different metropolitan areas, we think it is unlikely that large clusters of students from the same school participated, but we do not know.

This study was conducted with children enrolled in second grade. It is possible that the relationship between working memory processes and word learning processes changes over the course of development; therefore, findings may not generalize to younger or older students. The word learning tasks in this study assessed the triggering and configuration stages of word learning when children first encountered new words but did not assess the engagement stage of word learning when the representations of words are solidified in memory with experience. The relationship between working memory processes as a whole and the engagement stage of word learning likely differ from earlier word learning stages.

Conclusions

Using established structural equation models of working memory and word learning in young children, we found that working memory was a significant predictor of word learning over and above the contributions of expressive vocabulary and nonverbal IQ. Results suggest that a comprehensive working memory assessment could contribute important diagnostic information to inform sources of word learning difficulties. Studies are needed to determine whether tailoring instruction based on a child's working memory profile could increase learning.

Acknowledgments

This work was funded by National Institute on Deafness and Other Communication Disorders Grant R01 DC010784, awarded to Shelley I. Gray. We sincerely thank the children and their families who participated in this research and the following school districts and educational organizations from Arizona, Nebraska, New Hampshire, and Massachusetts that partnered with us to recruit participants: Academy of Math and Science, Alhambra School District, Arizona Museum for Youth, Chandler Unified School District, Desert Springs Academy, Foothills Academy, Carroll School, Carver Public Schools, City of Tempe Kid Zone Enrichment Program, Concordia East Valley Children's Theatre Boy's and Girl's Clubs, Faith Lutheran School, Gilbert Public Schools, Glendale Elementary School District, Imagine Schools, Isaac School District, KIDCO, La Paloma Academy, Lincoln Christian School, Lincoln Public Schools, Maricopa Unified School District, Melrose Public Schools, Mesa Public Schools, Nogales Unified School District, Saint John School, Saint Joseph School, Saint Mary of the Assumption School, Saint Patrick School, Saint Raphael School, Saint Teresa School, San Tan Charter School, Satori, Stoneham Public Schools, and Sunnyside Unified. We also thank Yi Feng and Greg Hancock for clarifying comments regarding their approach to incremental validity.

References

Adams, A., & Gathercole, S. E. (1996). Phonological working memory and spoken language development in young children. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49*(1), 216–233. https://doi.org/10.1080/713755610

Adams, A., & Gathercole, S. E. (2000). Limitations in working memory: Implications for language development. *International Journal of Language & Communication Disorders, 35,* 95–116. https://doi.org/10.1080/136828200247278

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106,* 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Development, 80,* 606–621. https://doi.org/10.1111/j.1467-8624.2009.01282.x

Alt, M., Hogan, T., Green, S., Gray, S., Cabbage, K., & Cowan, N. (2017). Word learning deficits in children with dyslexia. *Journal of Speech, Language, and Hearing Research, 60*(4), 1012–1028. https://doi.org/10.1044/2016_JSLHR-L-16-0036

Andrade, J., & Baddeley, A. (2011). The contribution of phonological short-term memory to artificial grammar learning. *Quarterly Journal of Experimental Psychology, 64*(5), 960–974. https://doi.org/10.1080/17470218.2010.533440

Apter, B. J. B. (2012). Do computerised training programmes designed to improve working memory work? *Educational Psychology in Practice, 28,* 257–272. https://doi.org/10.1080/02667363.2012.712915

Archibald, L. M. D., & Joanisse, M. F. (2013). Domain-specific and domain-general constraints on word and sequence learning. *Memory and Cognition, 41,* 268–280. https://doi.org/10.3758/s13421-012-0259-4

Baddeley, A. D. (1986). *Working memory*. Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Bentler, P. M. (1990). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bentler, P. M. (2006). *EQS 6 structural equations program manual*. Multivariate Software.

Bentler, P. M. (2007). Covariance structure models for maximal reliability of unit-weighted composites. In S. Lee (Ed.), *Handbook of computing and statistics with applications* (Vol. 1, pp. 1–19). Elsevier. https://doi.org/10.1016/S1871-0301(06)01001-8

Berninger, V. W., Abbott, R. D., Swanson, H. L., Lovitt, D., Trivedi, P., Lin, S. J., Gould, L., Youngstrom, M., Shimada, S., & Amtmann, D. (2010). Relationship of word- and sentence-level working memory to reading and writing in second, fourth, and sixth grade. *Language, Speech, and Hearing Services in Schools, 41*(2), 179–193. https://doi.org/10.1044/0161-1461(2009/08-0002)

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley. https://doi.org/10.1002/9781118619179

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Cabbage, K. L., Brinkley, S., Gray, S., Alt, M., Cowan, N., Green, S., Kuo, T., & Hogan, T. (2017). Assessing working memory in children: The Comprehensive Assessment Battery for Children–Working Memory (CABC-WM). *Journal of Visualized Experiments, 124,* Article e55121. https://doi.org/10.3791/55121

Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: The Dynamic Assessment of Word Learning (DAWL). *International Journal of Language & Communication Disorders, 48,* 565–581. https://doi.org/10.1111/1460-6984.12033

Carrow-Woolfolk, E. (2014). *TACL-4: Test for Auditory Comprehension of Language–Fourth Edition*. Pro-Ed. https://www.proedinc.com/Products/12700/tacl4-test-for-auditory-comprehension-of-languagefourth-edition.aspx

Chalmers, K. A., & Freeman, E. E. (2018). A comparison of single and multi-test working memory assessments in predicting academic achievement in children. *The Journal of Psychology, 152*(8), 613–629. https://doi.org/10.1080/00223980.2018.1491469

Chow, J. C., Ekholm, E., & Bae, C. L. (2021). Relative contribution of verbal working memory and attention to child language. *Assessment for Effective Intervention, 47*(1), 3–13. https://doi.org/10.1177/1534508420946361

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Erlbaum.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87–185. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2005). *Working memory capacity*. Psychology Press.

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26,* 197–223. https://doi.org/10.1007/s10648-013-9246-y

Cowan, N. (2016). *Working memory capacity: Classic edition*. Routledge. https://doi.org/10.4324/9781315625560

Cowan, N. (2019). Short-term memory based on activated long-term memory: A review in response to Norris (2017). *Psychological Bulletin, 145,* 822–847. https://doi.org/10.1037/bul0000199

Cunningham, J., & Sood, K. (2018). How effective are working memory training interventions at improving maths in schools: A study into the efficacy of working memory training in children aged 9 and 10 in a junior school? *Education 3–13, 46*(2), 174–187. https://doi.org/10.1080/03004279.2016.1210192

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422–433. https://doi.org/10.3758/BF03214546

Fazio, L. K., & Marsh, E. J. (2019). Retrieval-based learning in children. *Current Directions in Psychological Science, 28,* 111–116. https://doi.org/10.1177/0963721418806673

Feng, Y., & Hancock, G. R. (2021). Model-based incremental validity. *Psychological Methods.* https://doi.org/10.1037/met0000342

French, L. M., & Obrien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*(3), 463–487. https://doi.org/10.1017/S014271640808021

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*(4), 439–454. https://doi.org/10.1111/j.2044-8295.1990.tb02371.x

Gathercole, S. E., Brown, L., & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment levels. *Educational and Child Psychology, 20*(3), 109–122.

Giofre, D., Borella, E., & Mammarella, I. C. (2017). The relationship between intelligence, working memory, academic self-esteem, and academic achievement. *Journal of Cognitive Psychology, 29,* 731–747. https://doi.org/10.1080/20445911.2017.1310110

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation–Second Edition*. American Guidance Service Inc.

Gray, S., Alt, M., Hogan, T. P., Green, S., & Cowan, N. (n.d.). *Comprehensive assessment battery for children – Working memory* [Unpublished measure under development].

Gray, S., Fox, A., Green, S., Alt, M., Hogan, T., Petscher, Y., & Cowan, N. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research, 62*(6), 1839–1858. https://doi.org/10.1044/2019_JSLHR-L-18-0148

Gray, S., Green, S., Alt, M., Hogan, T., Kuo, T., Brinkley, S., & Cowan, N. (2017). The structure of working memory in young children and its relation to intelligence. *Journal of Memory and Language, 92,* 183–201. https://doi.org/10.1016/j.jml.2016.06.004

Gray, S., Lancaster, H., Alt, M., Hogan, T. P., Green, S., Levy, R., & Cowan, N. (2020). The structure of word learning in young school-age children. *Journal of Speech, Language, and Hearing Research, 63*(5), 1446–1466. https://doi.org/10.1044/2020_JSLHR-19-00186

Green, S. B., Yang, Y., Alt, M., Brinkley, S., Gray, S., Hogan, T., & Cowan, N. (2016). Use of internal consistency coefficients for estimating reliability of experimental task scores. *Psychonomic Bulletin Review, 23,* 750–763. https://doi.org/10.3758/s13423-015-0968-3

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future—A festschrift in honor of Karl Jöreskog* (pp. 195–216). Scientific Software International.

Holmes, J., & Gathercole, S. E. (2014). Taking working memory training from the laboratory into schools. *Educational Psychology, 34,* 440–450. https://doi.org/10.1080/01443410.2013.797338

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Memory and Language, 63,* 100–116. https://doi.org/10.1016/j.jml.2010.02.003

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jaroslawska, A. J., Gathercole, S. E., Logie, M. R., & Holmes, J. (2016). Following instructions in a virtual school: Does working memory play a role? *Memory & Cognition, 44,* 580–589. https://doi.org/10.3758/s13421-015-0579-2

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. E. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*. AGS.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford.

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology, 55*(4), 306–353. https://doi.org/10.1016/j.cogpsych

Leonard, L. B., & Deevy, P. (2020). Retrieval practice and word learning in children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research, 63*(10), 3252–3262. https://doi.org/10.1044/2020_JSLHR-20-00006

Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology, 18,* 62–78. https://doi.org/10.1080/09297049.2011.575772

Maehler, C., & Schuchardt, K. (2016). Working memory in children with specific learning disorders and/or attention deficits. *Learning and Individual Differences, 49,* 341–347. https://doi.org/10.1016/j.lindif.2016.05.007

Marini, A., Gentili, C., Molteni, M., & Fabbro, F. (2014). Differential verbal working memory effects on linguistic production in children with Specific Language Impairment. *Research in Developmental Disabilities, 35*(12), 3534–3542. https://doi.org/10.1016/j.ridd.2014.08.031

Masoura, E. V., & Gathercole, S. E. (2005). Contrasting contributions of phonological short-term memory and long-term knowledge to vocabulary learning in a foreign language. *Memory, 13*(3−4), 422–429. https://doi.org/10.1080/09658210344000323

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49,* 270–291. https://doi.org/10.1037/a0028228

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working memory and Specific Language Impairment: An update on the relation and perspectives on assessment and treatment. *American Journal of Speech-Language Pathology, 19*(1), 78–94. https://doi.org/10.1044/1058-0360(2009/09-0028)

Morra, S., & Camba, R. (2009). Vocabulary learning in primary school children: Working memory and long-term memory components. *Journal of Experimental Child Psychology, 104*(2), 156–178. https://doi.org/10.1016/j.jecp.2009.03.007

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Norris, D., Page, M. P. A., & Hall, J. (2018). Learning nonwords: The Hebb repetition effect as a model of word learning. *Memory, 26,* 852–857. https://doi.org/10.1080/09658211.2017.1416639

Passolunghi, M. C., Vercelloni, B., & Schadee, H. (2007). The precursors of mathematics learning: Working memory, phonological ability and numerical competence. *Cognitive Development, 22,* 165–184. https://doi.org/10.1016/j.cogdev.2006.09.001

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., Dardick, W., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin, 144,* 48–76. https://doi.org/10.1037/bul0000124

Randall, L., & Tyldesley, K. (2016). Evaluating the impact of working memory training programmes on children—A systematic review. *Educational and Child Psychology, 33,* 34–50.

Raykov, T. (2004). Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. *Behavior Therapy, 35,* 299–331. https://doi.org/10.1016/S0005-7894(04)80041-8

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Kane, M. J., & Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142,* 359–379. https://doi.org/10.1037/a0029082

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58*(2), 345–359. https://doi.org/10.1044/2015_JSLHR-L-14-0150

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rowe, M. L., Raudenbush, S. W., & Goldin-Meadow, S. (2012). The pace of vocabulary growth helps predict later vocabulary skill. *Child Development, 83,* 508–525. https://doi.org/10.1111/j.1467-8624.2011.01710.x

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66,* 507–514. https://doi.org/10.1007/BF02296192

Schrank, F. A., McGrew, K. S., & Mather, N. (2014). *Woodcock-Johnson Test of Cognitive Abilities–Fourth Edition*. Houghton Mifflin Harcourt.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition (CELF-4)*. AGS.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138,* 628–654. https://doi.org/10.1037/a0027473

Singer, B. D., & Bashir, A. S. (2018). Wait...what??? Guiding intervention principles for students with verbal working memory limitations. *Language, Speech, and Hearing Services in Schools, 49*(3), 449–462. https://doi.org/10.1044/2018_LSHSS-17-0101

Söderqvist, S., & Nutley, S. B. (2015). Working memory training is associated with long term attainments in math and reading.

Frontiers in Psychology, 6, 1711. https://doi.org/10.3389/fpsyg.2015.01711

Steiger, J. H. (2016). Notes on the Steiger–Lind (1980) handout. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(6), 777–781. https://doi.org/10.1080/10705511.2016.1217487

Steiger, J. H., & Lind, J. C. (1980). *Statistically-based tests for the number of common factors* [Paper presentation]. Annual Meeting of the Psychometric Society, Iowa City, IA, United States.

Swanson, H. L., & Berninger, V. W. (1996). Individual differences in children's working memory and writing skill. *Journal of Experimental Child Psychology, 63,* 358–385. https://doi.org/10.1006/jecp.1996.0054

Swanson, J., Deutsch, C., Cantwell, D., Posner, M., Kennedy, J., Barr, C., Moyzis, R., Schuck, S., Flodman, P., & Spence, A. (2001). Genes and attention-deficit hyperactivity disorder. *Clinical Neuroscience Research, 1*(3), 207–216. https://doi.org/10.1016/S1566-2772(01)00007-X

Sweller, J. (2011). Cognitive load theory. *Psychology of Learning and Motivation, 55,* 37–76. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Torgesen, J., Wagner, R., & Rashotte, C. (2012). *Test of Word Reading Efficiency–Second Edition (TOWRE-2)*. AGS.

van Haaften, L., Diepeveen, S., van den Engel-Hoek, L., Jonker, M., de Swart, B., & Maassen, B. (2019). The psychometric evaluation of a speech production test battery for children: The reliability and validity of the Computer Articulation Instrument. *Journal of Speech, Language, and Hearing Research, 62*(7), 2141–2170. https://doi.org/10.1044/2018_JSLHR-S-18-0274

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141,* 65–82. https://doi.org/10.1016/j.jecp.2015.06.015

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. NCS Pearson.

Williams, K. T. (2007). *Expressive Vocabulary Test–Second Edition (EVT-2)*. Pearson Assessments. https://doi.org/10.1037/t15094-000

Woodcock, R. W. (2011). *Woodcock Reading Mastery Tests–Third Edition manual*. Pearson Clinical Assessment Solutions.